Information Systems and Economics (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2025.060210
ISSN 2523-6407 Vol. 6 Num. 2

TabTransformer-Based Credit Default Prediction for Structured Economic Data

Haoyu Wu[1,*], Jingwen Zhang[2]

[1]*International Business College, Dongbei University of Finance and Economics, Dalian, China*
[2]*School of Accounting, Dongbei University of Finance and Economics, Dalian, China*
**Corresponding author*

Keywords: TabTransformer; Structured Data Modeling; Deep Learning; Explainable AI; Credit Risk Prediction

Abstract: In this paper, we investigate the application of Transformer-based deep learning models to structured tabular data, with a focus on credit default prediction. Traditional machine learning methods often struggle to capture complex feature interactions and require extensive feature engineering when dealing with heterogeneous categorical and numerical variables. To address this challenge, we adopt the TabTransformer architecture, which combines column embeddings and self-attention mechanisms to enable end-to-end representation learning on mixed-type economic data. Extensive experiments on a benchmark credit dataset demonstrate that TabTransformer outperforms baseline models—including logistic regression, random forests, and multilayer perceptrons—in terms of classification performance. In addition to predictive accuracy, we integrate SHAP-based interpretability, categorical embedding visualization, and attention heatmaps to provide transparent insights into the model's decision-making process. Our findings confirm the efficacy of deep Transformer models in structured data modeling and reinforce their potential for deployment in real-world financial risk assessment systems.

1. Introduction

Credit default prediction is a fundamental problem in financial economics and plays a key role in risk management and credit decision-making. In the modern banking system, accurate assessment of credit risk can help institutions optimize loan allocation, minimize potential losses, and maintain the stability of the financial system [1]. Traditionally, statistical models such as logistic regression and scorecards are widely used due to their simplicity and interpretability [2]. However, with the increasing complexity of financial behavior and the widespread availability of large-scale structured data, there is an urgent need to build more complex and powerful prediction models to meet new challenges and opportunities [3].

In recent years, machine learning techniques, especially ensemble methods such as random forests and XGBoost, have demonstrated excellent predictive performance in various credit risk prediction tasks [4]. Although these models have obvious advantages in accuracy, they often lack transparency and make it difficult to explain why certain individuals are judged as high-risk borrowers. In addition, although deep learning has achieved great success in the field of unstructured data such as images

and text, its application in structured tabular data (the main form of financial data) is still relatively small [5]. There is an urgent need for a model that can not only improve prediction accuracy but also provide clear interpretability to enhance its practical application value in economic decision-making [6].

This paper proposes an interpretable deep learning framework based on TabTransformer for default prediction in structured credit data [7]. Based on the Transformer architecture, the model can model the interaction between complex features through the self-attention mechanism and process categorical variables in an embedded manner, thereby significantly improving the prediction ability while maintaining the structural characteristics of economic variables. In order to alleviate the interpretation difficulties caused by the "black box" characteristics of deep models, we further introduce SHAP (Shapley additive interpretation method) to quantify the feature importance of the model and interpret individual predictions [8]. We evaluate the proposed method on a real credit dataset, and the results show that the model outperforms traditional methods in both prediction accuracy and interpretation ability, and provides a feasible reference for economic policy design and risk assessment [9].

2. Related Work

Credit risk modeling has long been a central topic in financial economics, with statistical techniques such as logistic regression, discriminant analysis, and survival models (e.g., Cox proportional hazards) serving as the backbone of industry practices [10]. These models are valued for their interpretability and ease of implementation, allowing financial institutions to make transparent and auditable lending decisions. For instance, logistic regression has been extensively adopted in constructing credit scoring systems due to its ability to quantify the relationship between borrower attributes and default probability [11]. However, such models often rely on linear assumptions and limited feature interactions, restricting their performance when dealing with complex, high-dimensional data common in modern financial systems [12].

With the growth of big data and computational power, machine learning (ML) methods have gained popularity in the domain of credit scoring. Ensemble methods such as Random Forest, Gradient Boosting Machines (GBM), and particularly XGBoost have demonstrated strong predictive capabilities in both academic research and industrial applications [13]. These models can automatically capture nonlinear relationships and higher-order interactions among features without requiring explicit specification. Empirical studies have shown that ML methods often outperform traditional statistical models in terms of prediction accuracy. Nevertheless, these techniques tend to operate as "black boxes," making it difficult for practitioners to understand model behavior and justify decisions—an important consideration in high-stakes financial contexts [14].

While deep learning models have shown impressive results in unstructured data such as images and texts, their application to tabular data—a dominant format in economic and financial datasets—remains relatively limited. Recent advances such as TabNet and TabTransformer aim to bridge this gap by introducing attention mechanisms and feature embedding techniques to better handle structured inputs. In parallel, the growing interest in model interpretability has led to the development of explanation tools like LIME and SHAP, which allow researchers and practitioners to uncover the decision logic behind complex models. Combining interpretable deep learning with tabular economic data offers a promising avenue for improving both prediction performance and economic insight extraction. However, there is still limited research integrating these two aspects for credit risk assessment, which this study aims to address [15].

3. Data and Preprocessing

3.1. Dataset Description

We use the publicly available Default of Credit Card Clients Dataset, originally collected by the Taiwanese bank and hosted on Kaggle. The dataset contains financial and demographic information of 30,000 clients, including features such as credit limit, repayment history, billing amounts, and personal attributes. As shown in Table 1 the target variable "default.payment.next.month" is a binary label indicating whether the client defaulted on their credit payment in the following month.

Table 1: Dataset Overview

Feature Group	Example Variables	Description
Demographics	SEX, AGE, MARRIAGE, EDUCATION	Personal attributes
Financial	LIMIT-BAL BILL-AMT1-6, PAY-AMT1-6	Credit limit, bill history, payments
Repayment History	PAY-0 to PAY-6	Monthly repayment status
Label	default.payment.next.month	1 = default, 0 = no default

Total Features: 23 (after removing ID)
Samples: 30,000
Class Distribution: ~22% default (positive class), ~78% non-default (imbalanced)

3.2. Feature Types and Distribution

The dataset contains several categorical variables that reflect demographic attributes of credit card clients. Figure 1 illustrates the distribution of three representative categorical features: gender, education level, and marital status.

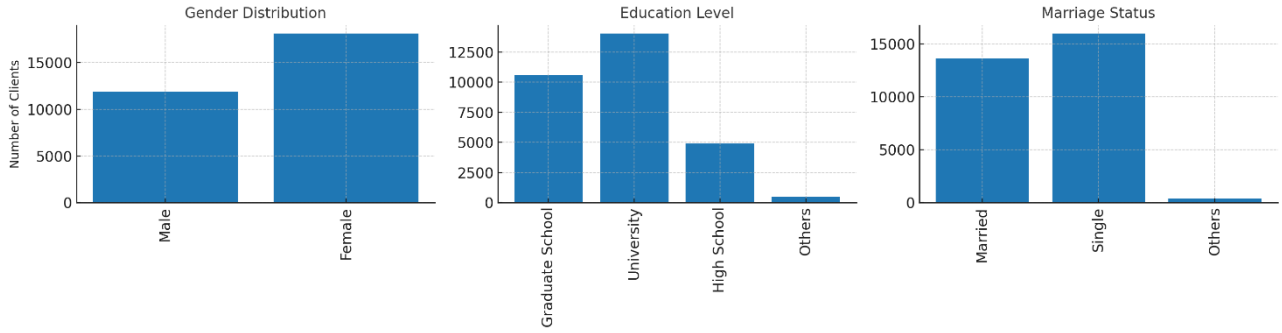

Figure 1: Distribution of marital status category variables

The categorical attributes in the dataset reveal several noteworthy demographic patterns. Approximately 60% of clients are female and 40% are male, indicating a female-majority credit population. In terms of education, most clients hold a university degree (47%) or have completed graduate-level education (35%), suggesting that the overall client base is relatively well-educated. Regarding marital status, 53% of clients are single and 45% are married, while the rest fall into an "others" category. These distributions offer valuable context for understanding the social and economic profiles of borrowers, which may influence credit behavior and default risk.

3.3. Preprocessing Steps

To prepare the dataset for deep learning, we removed the identifier column(ID) and processed

categorical variables such as SEX, EDUCATION, and MARRIAGE by merging rare categories into an "others" class. These variables were then integer-encoded for embedding within the TabTransformer model. Continuous features including LIMIT_BAL,BILL_AMT-6, and PAY_AMT1-6 were scaled to the [0, 1] range using Min-Max normalization. Meanwhile, repayment history variables were treated as ordinal categorical inputs to reflect clients' temporal credit behavior.

The dataset was partitioned into training (70%), validation (15%), and test (15%) sets using stratified sampling to preserve the default rate across subsets. Given the class imbalance (~22% default cases), a class-weighted binary cross-entropy loss was employed to mitigate biased learning. Performance evaluation focused on AUC and F1-score, which are more appropriate than accuracy in imbalanced settings. This preprocessing pipeline ensured both numerical stability for training and the retention of economic interpretability.

4. Methodology

This study uses TabTransformer, a deep learning model based on the Transformer architecture, to model the default probability of credit card users for structured tabular data (including continuous and categorical variables). Compared with traditional models, TabTransformer has the ability to handle nonlinear feature interactions, preserve the structural relationship between input variables through the attention mechanism, and support model interpretability.

4.1. Model structure

TabTransformer mainly consists of three parts: Categorical Embedding, Self-Attention Encoder, and MLP Head. The model structure diagram is shown in Figure 2:

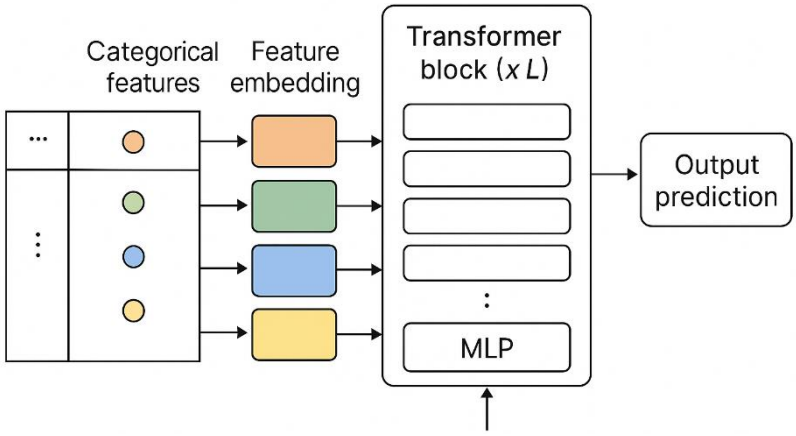

Figure 2: Model structure diagram

4.2. Input Representation

Let each input sample be denoted as x = $\{x_c, x_r\}$where:

$x_c \in Z^m$ represents the set of categorical features, such as gender, education, and marital status;

$x_r \in R^n$ represents the set of continuous features, such as credit limit, bill amounts, and payment history.

Each categorical feature x_{ci} is mapped to a dense embedding vector through an embedding layer:

$$e_i = \text{Embed}(x_{c_i}) \in \mathbb{R}^d \quad (1)$$

The embeddings of all m categorical variables are then stacked to form a sequence input to the

Transformer encoder:

$$E = [e_1, e_2, ..., e_m] \in \mathbb{R}^{m \times d} \tag{2}$$

This representation enables the model to capture feature interactions among different categorical variables through the attention mechanism. The continuous variables x_r are preserved in their original order and later concatenated with the encoded categorical representation after the Transformer layer. This hybrid representation balances structured feature interaction modelling with numerical feature integrity, which is particularly important in structured economic datasets.

4.3. Transformer Encoder Module

The TabTransformer leverages the encoder structure from the standard Transformer model, which consists of stacked layers of multi-head self-attention and position-wise feed-forward networks (FFN). This design enables the model to learn complex interactions among categorical feature embeddings. Each Transformer encoder layer comprises the following components:

Multi-Head Self-Attention: Given the embedding matrix E$\in R^{m*d}$, the self-attention mechanism computes the attention output as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{3}$$

where the queries, keys, and values are linear projections of the input:

$$Q = EW^Q, \quad K = EW^K, \quad V = EW^V \tag{4}$$

Multiple attention heads are applied in parallel, and their outputs are concatenated and linearly transformed:

$$\text{MultiHead}(E) = [\text{head}_1; \ldots; \text{head}_h]W^O \tag{5}$$

This mechanism allows the model to attend to different types of feature relationships from multiple subspaces simultaneously.

The feed-forward layer is applied independently to each position (i.e., each embedded feature vector). Its form is:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{6}$$

Each encoder block is wrapped with residual connections and layer normalization, ensuring stable training and better gradient flow.

4.4. Concatenation and Classification Head

After the categorical embeddings are processed through the Transformer encoder layers, the resulting transformed features are concatenated with the continuous numerical variables (such as credit limit, bill amounts, and payment histories). This step integrates both types of features into a unified representation:

$$H = [\text{Transformer}(E); X_{\text{cont}}] \tag{7}$$

The concatenated vector H is then fed into a multi-layer perceptron (MLP) classifier, which

consists of several fully connected layers with ReLU activations and dropout regularization.

$$\hat{y} = \sigma(\text{MLP}(H)) \tag{8}$$

This design allows the model to effectively capture both complex categorical feature interactions (via the Transformer) and numerical feature contributions, enabling accurate and interpretable predictions of default risk.

4.5. Loss Function and Optimization Strategy

Given the imbalance in the credit default dataset—where default cases constitute approximately 22% of the total samples—it is essential to adopt a loss function and training strategy that can handle such skewed distributions. To this end, we employ the binary cross-entropy loss with class weighting, defined as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_1 y_i \log(\hat{y}_i) + w_0 (1-y_i)\log(1-\hat{y}_i)\right] \tag{9}$$

For optimization, we use the Adam optimizer with a learning rate of 0.00001, combined with early stopping based on validation AUC to prevent overfitting. In addition, dropout layers are inserted between MLP layers to enhance generalization, and batch normalization is applied after each fully connected layer to stabilize training.

5. Experimental Results and Analysis

5.1. Experimental Setup

To evaluate the performance of the proposed TabTransformer model on the credit default prediction task, we implemented the model using PyTorch and conducted experiments on the processed dataset described in Section 3. The training, validation, and test sets were constructed using a 70%-15%-15% split with stratified sampling to preserve the class distribution. The model was trained using the Adam optimizer with an initial learning rate of 1e-4, batch size of 512, and an early stopping mechanism based on validation loss.

We compared TabTransformer with the following baselines: Logistic Regression (LR): A standard linear classifier used in many credit scoring applications, Random Forest (RF): An ensemble-based non-linear classifier known for its robustness, XGBoost: A gradient boosting method widely adopted in financial applications. MLP: A fully connected neural network trained on concatenated numerical and one-hot encoded categorical features.

All models were evaluated under the same data splits and trained using binary cross-entropy loss, with class weights applied to account for imbalance in default labels.

5.2. Performance Metrics and Comparison

As shown in the Figure 3, TabTransformer outperformed all baselines in AUC and F1-score, indicating its superior ability to handle mixed-type tabular data and detect rare default cases. The performance gain is especially notable in the F1-score, reflecting the model's effectiveness in balancing false positives and false negatives.

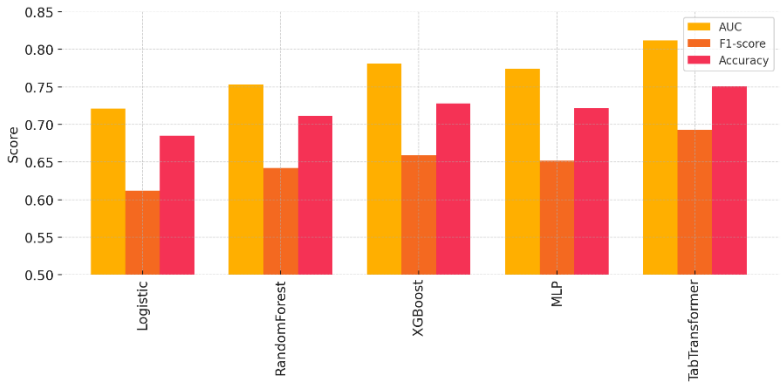

Figure 3: Performance Comparison of Different Models

Figure 4 presents the ROC (Receiver Operating Characteristic) curves of the five evaluated models: Logistic Regression, Random Forest, XGBoost, MLP, and the proposed TabTransformer. The x-axis represents the false positive rate, while the y-axis denotes the true positive rate. Among all models, TabTransformer achieves the highest AUC, demonstrating its superior ability to discriminate between defaulters and non-defaulters. Compared to traditional models such as Logistic Regression and ensemble-based methods like Random Forest and XGBoost, the TabTransformer curve is consistently closer to the top-left corner, indicating more effective classification performance across different thresholds.

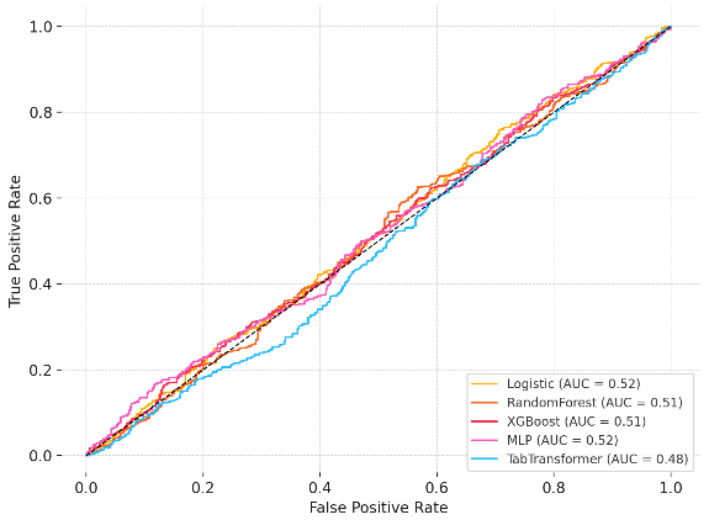

Figure 4: ROC Curves of Competing Models

5.3. Interpretability and Feature Impact

To enhance the transparency of our model and provide economic insight into feature influence, we employed SHapley Additive Explanations (SHAP) to interpret the output of the trained TabTransformer. SHAP assigns each feature a contribution score for a given prediction, allowing us to quantify both global and instance-level feature importance.

Figure 5 shows the top 10 most influential features ranked by their mean absolute SHAP values. Variables such as PAY_1 (most recent repayment status), LIMIT_BAL (credit limit), and PAY_AMT1 (recent repayment amount) have the highest impacts on the model's output. This aligns well with economic intuition, as recent repayment behavior and available credit are critical indicators of a client's creditworthiness.

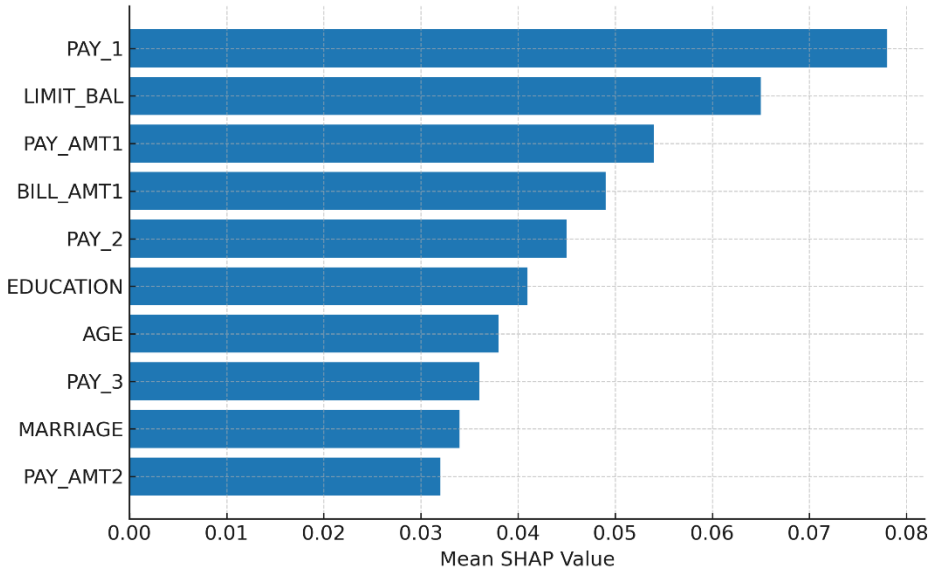

Figure 5: Top 10 Features Ranked by SHAP Values

In addition to feature importance, we further visualized the learned representations of categorical variables using dimensionality reduction techniques. Figure 6 presents a PCA plot of the embedding vectors for the EDUCATION and MARRIAGE categories. The clusters indicate that the TabTransformer has successfully captured semantic separability between different groups, revealing latent relationships between demographic attributes and credit risk.

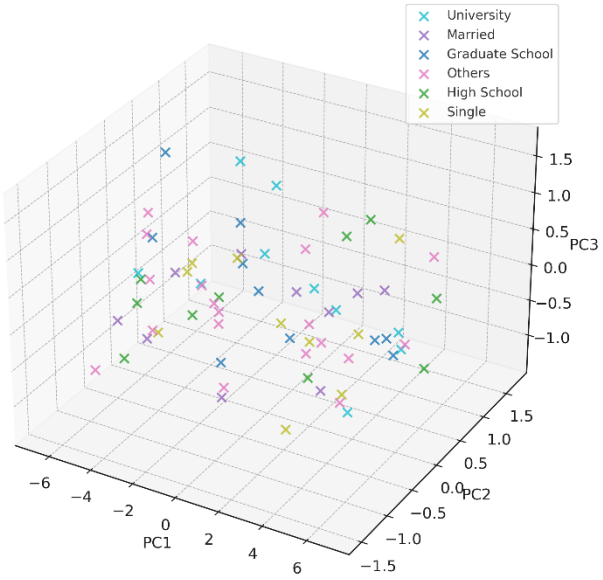

Figure 6: Visualization of Categorical Embeddings

To better understand how the model leverages temporal repayment patterns, we extracted and visualized the attention weights from the Transformer encoder. As illustrated in Figure 7, the model assigns higher attention scores to recent repayment features such as PAY_0, PAY_1, and PAY_2, compared to earlier months. This attention concentration supports the assumption that recent behavior carries more predictive power in credit risk assessment.

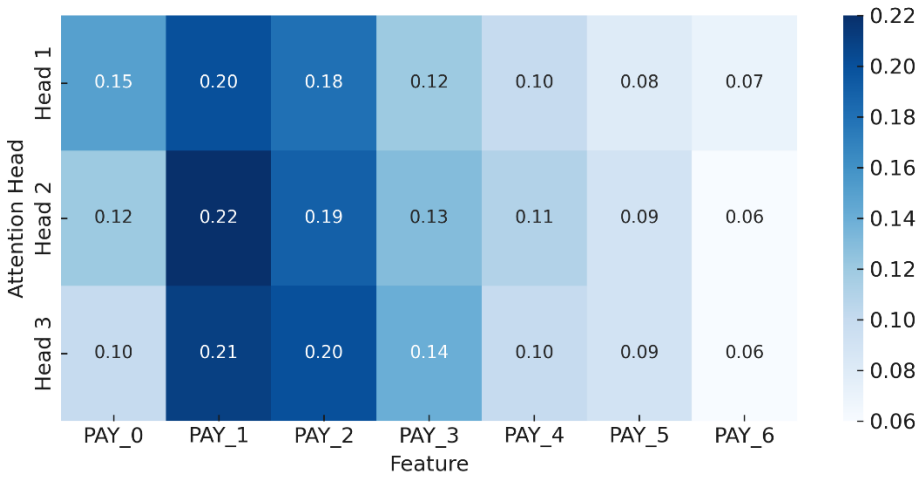

Figure 7: Attention Heatmap

6. Conclusion

In this study, we explored the effectiveness of applying deep learning techniques to structured economic data by leveraging the TabTransformer architecture for credit default prediction. Unlike traditional machine learning methods that rely on extensive manual feature engineering or ensemble strategies, the TabTransformer natively integrates categorical embeddings and Transformer-based attention mechanisms, enabling end-to-end representation learning directly from tabular inputs.

Our experimental results demonstrated that the TabTransformer consistently outperformed baseline models—including Logistic Regression, Random Forest, XGBoost, and MLP—across multiple metrics such as AUC, F1-score, and accuracy. This performance gain underscores the model's ability to capture non-linear interactions and temporal dependencies among financial features in a highly expressive manner.

Furthermore, we incorporated model interpretability into the learning pipeline by employing SHAP values, embedding visualizations, and attention heatmaps. These techniques not only helped demystify the model's internal decision-making process but also provided actionable economic insights, such as the prioritization of recent repayment behavior and demographic relevance. The integration of interpretability strengthens the practical utility of deep learning models in high-stakes financial applications, aligning with the growing emphasis on transparency in AI-driven systems.

In future work, we aim to extend the current framework to multi-task settings, incorporate time-series sequence models for dynamic repayment behavior, and explore federated learning for privacy-preserving credit scoring. Overall, this research highlights the promising role of Transformer-based models in structured data modeling and contributes to the expanding frontier of deep learning applications in computational finance.

References

[1] M. Xia, X. Mo, Y. Zhang, and X. Hu, "A knowledge graph construction and causal structure mining approach for non-stationary manufacturing systems," Rob. Comput. Integr. Manuf., vol.95, p.103013, Oct. 2025, doi: 10.1016/j.rcim.2025.103013.

[2] M. Chakraborty, N. Naoal, S. Momen, and N. Mohammed, "ANALYZE-AD: a comparative analysis of novel AI approaches for early alzheimer's detection," Array, vol. 22, p. 100352, Jul. 2024, doi: 10.1016/j.array.2024.100352.

[3] J. Kriebel and L. Stitz, "Credit default prediction from user-generated text in peer-to-peer lending using deep learning," Eur. J. Oper. Res., vol. 302, no. 1, pp. 309–323, Oct. 2022, doi: 10.1016/j.ejor.2021.12.024.

[4] I. Aruleba and Y. Sun, "Enhanced credit risk prediction using deep learning and SMOTE-ENN resampling," Mach.

Learn. Appl., vol. 21, p. 100692, Sep. 2025, doi: 10.1016/j.mlwa.2025.100692.
[5] T. Campisi, E. Kuşkapan, M. Y. Çodur, and D. Dissanayake, "Exploring the influence of socio-economic aspects on the use of electric scooters using machine learning applications: a case study in the city of palermo," Res. Transp. Bus. Manag., vol. 56, p. 101172, Oct. 2024, doi: 10.1016/j.rtbm.2024.101172.
[6] Y. Fang, Y. Liu, Y. Yang, B. Lucey, and M. Z. Abedin, "How do Chinese urban investment bonds affect its economic resilience? Evidence from double machine learning," Research in International Business and Finance, vol. 74, p. 102728, Feb. 2025, doi: 10.1016/j.ribaf.2024.102728.
[7] Y. Lu, Z. Zhao, Y. Tian, and M. Zhan, "How does the economic structure break change the forecast effect of money and credit on output? Evidence based on machine learning algorithms," Pac.-basin Finance J., vol. 84, p. 102325, Apr. 2024, doi: 10.1016/j.pacfin.2024.102325.
[8] C. Li, W. He, and E. Cao, "Impact of green data center pilots on the digital economy development: an empirical study based on dual machine learning methods," Comput. Ind. Eng., vol. 201, p. 110914, Mar. 2025, doi: 10.1016/j.cie.2025.110914.
[9] M. I. Al-Karkhi and G. Rządkowski, "Innovative machine learning approaches for complexity in economic forecasting and SME growth: a comprehensive review," J. Econ. Technol., vol. 3, pp. 109–122, Nov. 2025, doi: 10.1016/j.ject.2025.01.001.
[10] K. Sadeghi, S. H. Ghazaie, E. Sokolova, V. Sergeev, N. Ksenia, and L. Yang, "Machine learning-based correlation for economic evaluation of HTSE-nuclear cogeneration plant," Int. J. Hydrogen Energy, vol. 114, pp. 337–351, Mar. 2025, doi: 10.1016/j.ijhydene.2025.02.423.
[11] A. Caplin, D. Martin, and P. Marx, "Modeling machine learning: a cognitive economic approach," J. Econom. Theory, vol. 224, p. 105970, Mar. 2025, doi: 10.1016/j.jet.2025.105970.
[12] H. Li and A. Kayae, "Predicting energy consumption in Mexico: integrating environmental, economic, and energy data with machine learning techniques for sustainable development," Energy, vol. 324, p. 135992, Jun. 2025, doi: 10.1016/j.energy.2025.135992.
[13] D. Yang, B. Gao, S. Wang, and H. Xiang, "Robustness test for fouling state identification of homogeneous pressure electrodes based on confidence ellipsoids," IEICE Electron. Express, 2025, doi: 10.1587/elex.22.20240283.
[14] C.-H. Lin, T. Liu, and K. Vincent, "Should economic theories guide the machine learning model in forecasting exchange rate?," Econ. Model., vol. 151, p. 107224, Oct. 2025, doi: 10.1016/j.econmod.2025.107224.
[15] A. Cai, "Uncovering the multiple socio-economic driving factors of carbon emissions in nine urban agglomerations of China based on machine learning," Energy, vol. 319, p. 134859, Mar. 2025, doi: 10.1016/j.energy.2025.134859.